You are given the transcriptions of 3 news shows. Each show has an ID:

18246, 16859 and 12387

Transcriptions contain the full text of the news show plus the exact starting time and ending time of each word.

You are tasked with adding 4 columns to this dataframe.
Each row of this dataframe is 1 news story from the news show in the source_video_id column.
The column first_words contains the first 8 words of the story, and the column last_words contains the last 8 words of the story.

Your first task is to add three columns to the dataframe:
body: The full text of each story
start: starting time of the story
end: ending time of the story

(Obviously, you should extract these from the transcription files using the starting words and ending words of each story.)

Your second task is to predict the topics of each story. Add a column named topics to the dataframe which contains the list of predicted topics for each row.

For training, you can use this dataframe.

The lists in the topic column contain the topic ids of that story.
Your column should have the same format.

IMPORTANT: Good thinking and clean code matters much more than the results.